CLINICAL REVIEW ARTICLE

The principles of cost-effectiveness analysis and their application

Brenda K. Zierler, PhD,[a,b,d] and **Darryl T. Gray, MD, ScD,**[c,d,e,f] *Seattle, Wash*

The current healthcare environment requires the evaluation of both the costs and benefits of alternative interventions for a given clinical problem. Given the increased interest in the economic evaluation of healthcare interventions, this article briefly defines various forms of economic evaluations and describes some useful steps for conducting appraisals of cost-effectiveness analyses. Studies of competing methods of treatment of abdominal aortic aneurysms greater than 5 cm are used as a clinical example of interest to the readers of this Journal. Rather than actually conducting such an analysis with existing data, we describe the principles for conducting or reviewing an economic analysis with factitious data. (J Vasc Surg 2003;37:226-34.)

A first-year surgery resident describes a new patient seen in the clinic as a 75-year-old man with an asymptomatic 5-cm abdominal aortic aneurysm (AAA). The patient had an myocardial infarction at age 63 years but now has good exercise tolerance. The patient also has mild hypertension with a blood pressure of 165/88 mm Hg with two medications. The patient does not have diabetes, and his renal function is normal. The patient smokes one half of a pack of cigarettes per day. The resident asks whether an open surgical repair (OSR) rather than an endovascular repair (EVR) of the AAA would be more cost effective.

The attending physician, who is skeptical of the newer, less-invasive EVR method and of all of the associated follow-up studies that monitoring its outcome necessitates, says, "Yes, it is. The endografts cost about $12,000 to $15,000 at our hospital, and the cost of these grafts offsets the extended length of stay for OSR, which is about 9 days at our institution, compared with 3 days for EVR. But there are also many late problems with the endografts that require ongoing diagnostic studies and possibly additional interventions. So, you have to add the costs of identifying and treating possible leaks and another $16,000 for an open repair if the EVR fails. Our combined mortality and long-term morbidity rate of 2% for OSR is similar to EVR rates reported in the literature, so I am quite confident that OSR is more cost effective than EVR of AAAs."

One can use this clinical example to illustrate the principles of cost-effectiveness analysis as applied to a vascular surgery question. The attending physician's statements about which intervention would be more cost effective for the treatment of AAAs raise more questions than they answer, such as:

Has he conducted a formal economic comparison?
What assumptions are being made?
Is he quoting charges or costs?
What are the sources for the effectiveness data?
What were the primary outcomes of interest?
How accurate are the estimates for determining the combined mortality and morbidity rates for OSR of AAAs at his institution?
What is the time horizon for analyzing his data?
Is he quoting results for comparable populations of patients, such that differences in mortality and morbidity rates reflect treatment effects rather than patient characteristics?

Although OSR for AAA is a relatively safe procedure in some patient groups,[1] EVR was developed to avoid the risk of open abdominal surgery in patients who were poor operative candidates. EVR exemplifies new technology for which financial and clinical implications must be evaluated in the current environment.[2] The medical literature increasingly includes economic analyses[3] and guidelines for conducting, reporting, and reviewing such studies.[4-10] However, adherence to these guidelines in the cardiovascular surgery literature has been variable.[11]

From the Department of Biobehavioral Nursing and Health Systems, School of Nursing,[a] the Departments of Surgery[b] and Pediatrics,[c] School of Medicine, the Department of Health Services, School of Public Health and Community Medicine,[d] and the Center for Cost Outcomes Research,[e] University of Washington; and The Hope Heart Institute.[f]

Supported by NIH Grant NR00143 and American Heart Association Grant-in-Aid # 0050361N.

Competition of interest: nil.

Additional material for this article may be found online at www.mosby.com/jvs.

Reprint requests: Brenda K. Zierler, PhD, Department of Biobehavioral Nursing and Health Systems, Box 357266, 1959 NE Pacific St, Seattle, WA 98195-7266 (e-mail: brendaz@u.washington.edu).

0741-5214/2003/$30.00 + 0
doi:10.1067/mva.2003.59

A literature search identified 14 English-language articles that compared costs and outcomes of OSR and EVR of AAAs.[12-25] An informal review indicated substantial room for improvement in adherence to principles of economic analysis. For example, claims made in a number of articles about the cost effectiveness or cost benefit of one strategy or the other were not well supported by the accompanying data.

The goals of this article are to briefly define various forms of economic evaluations and to describe some useful steps for conducting an appraisal of a cost-effectiveness analysis, with a clinical example of interest to the readers of this Journal. Our intent is not to actually conduct such an analysis with existing data but to describe the principles for conducting or reviewing an economic analysis with realistic data.[12-25] The controversy about treatment of AAAs less than 5 cm or between 5 cm and 6 cm is beyond the scope of this discussion. For the purpose of this paper and as an academic exercise only, we have arbitrarily chosen 5 cm as the threshold for intervention in comparing OSR with EVR. Although watchful waiting may also be an appropriate alternative for some AAAs between 5 and 5.5 cm, we have not considered this as an alternative in our example. However, it is important to note that if a real cost-effectiveness analysis were being conducted, the size threshold for intervention and the methods used for determination of AAA size, along with alternative interventions for comparisons, would need to be stated explicitly.

OVERVIEW OF ECONOMIC EVALUATIONS: BASIC DEFINITIONS

For a given level of available resources, the goal of maximizing total aggregate health outcomes assumes that improvement in health status may be achieved through clinical interventions that have some measurable cost, provided their relative cost effectiveness is properly assessed. One value of formal cost analyses is that they force explicit consideration of the benefits to be anticipated from such interventions and of the costs incurred to achieve them. Various treatment alternatives can be compared in terms of cost and outcome estimates and value tradeoffs to be made in light of relevant ethical, political, and other considerations.

The most basic form of comparative cost analysis is a simple cost comparison, which merely asks how much two competing interventions cost. This first step may help determine whether or not further investigation is warranted. However, results of such an analysis cannot be meaningfully interpreted without information on the clinical value of the approaches also being considered; a procedure not worth doing is not worth doing well (ie, efficiently).[8] Thus, simple cost comparisons represent departure points rather than destinations.

Cost-effectiveness analysis and *cost-benefit analysis* are two approaches that simultaneously consider costs and outcomes of therapy. Confusion regarding the proper use of these terms is widespread, and studies that are labeled as cost-effectiveness analyses are often in fact cost-benefit analyses, and vice versa.[26] Cost-benefit and cost-effectiveness analyses are similar in that they both quantify the monetary cost of the clinical intervention. The major distinction between these two approaches is that cost-benefit analyses require both costs and benefits to be described in monetary units (generally currency) and cost-effectiveness analysis imposes no such requirement on the expression of benefits.[8] One attractive feature of cost-benefit analysis is that cost-benefit ratios for interventions with entirely different clinical objectives may be compared directly (eg, comparing the cost-benefit ratio of creating a new endovascular center with that of starting a new dialysis center). This is important when decisions on assigning available resources to various treatments need to be made. Not surprisingly, use of cost-benefit methods to express the benefits of interventions in monetary terms can be extremely complex.

Although cost-effectiveness analysis also measures costs in monetary units, the clinical outcomes are left in natural units rather than being converted to currency. These units can range from satisfying physiologic criteria (eg, absolute or relative reduction in) to achieving a stated clinical objective (eg, successful exclusion of an AAA) or to avoiding or prolonging the interval before some undesirable consequence (eg, rupture of a small AAA). The cost-effectiveness approach avoids valuing such outcomes in terms of currency but is restricted to comparisons of interventions with goals that can be expressed in the same units (eg, subsequent changes in expected mortality).[9]

Clinicians and policymakers implicitly assign monetary values to outcomes by choosing whether or not to provide specific clinical services. For example, the decision to undertake a therapeutic regimen that improves mean survival by 1 year for every $50,000 spent implies that a life-year is worth at least $50,000. If an available intervention with which 1 additional life-year can be saved at a cost of $500,000 is not undertaken, it may be inferred that the decision maker does not consider this outcome to be worth the required expenditure. However, such decisions are not always made consistently, in that some interventions with higher costs per life-year saved may be provided and others with lower costs per life-year saved are not.[8,9]

Cost-minimization analysis may be considered to be a special type of cost-effectiveness analysis in which one first establishes that the clinical outcomes of two interventions are equal. The focus then is placed on identification of the option with the lower cost. However, this requires that the equivalence of outcomes for the two approaches (ie, the absence of clinically and statistically significant differences) be proven and not just assumed.[8]

Cost-utility analysis is a more comprehensive approach in which the clinical outcomes are expressed as utilities, in the form of preference-weighted units such as the commonly used quality-adjusted life-years (QALYs; described subsequently), rather than only as unadjusted effectiveness measures such as crude survival. In economic comparisons, health outcomes are the end results of the interventions

performed with the goal of affecting the health status of the patient population or hypothetic cohort, as measured from the time of the intervention to the end of the observational period (or death).[27] Individuals or society may have a desire or preference for a given outcome, to which they may be said to assign a utility value, although sometimes the terms utility, preference, and value are used interchangeably. Several different methods (eg, standard gamble, time-trade off, rating scale) are used to determine the value that investigators assign to a given health state or health benefit. Preference-based methods involve judgements about the value placed on a particular health state and yield a single score that is scaled from 0.0 to 1.0 (where 0 usually means death and 1 usually means perfect health).[8,9,27]

The use of QALYs allows potential differences in survival and morbidity attainable with different interventions to be compared with a single measure that reflects tradeoffs between them. The quality weights must meet three criteria to be useful for generating QALYs. They must be: 1, anchored on two required points (usually perfect health and death); 2, based on preferences for the health states (eg, with greater values assigned to more preferred states); and 3, measured on an interval scale (eg, with weight of 1 as perfect health and weight of 0 as death).[9] Note that some might consider severe intractable pain or a debilitating stroke to be "worse than death" and use these outcomes as 0 anchors, with death assigned some value such as 0.1. In any case, quality adjustment weights for each health state are multiplied by the time in that state and then summed to calculate the number of quality adjusted life-years. Even if reliable data are available on crude life expectancy after the interventions of interest, the quality adjustment itself can be a subjective exercise.[8,9] More and more cost analyses now follow current recommendations that studies consider utilities or preferences for individual outcomes such as survival with versus without procedure complications or treatment side effects,[8,9,27] rather than just crude survival, for example. Finally, it should be noted that the term cost-effectiveness analysis is increasingly used to describe studies that are, strictly speaking, cost-utility analyses.[9]

Cost-consequence analysis is a less restrictive form of comparative cost analysis. In a cost-consequence analysis, one provides information about various costs (eg, provider costs, patient out-of-pocket costs, payer costs) and outcomes (eg, survival, complications, clinical success rates). However, the readers are left to decide which combination of costs and outcomes to combine to generate the cost-effectiveness ratios most relevant to their own perspectives.[28]

USE OF PUBLISHED GUIDELINES FOR PERFORMING ECONOMIC ANALYSES OF AAA REPAIR

With a checklist that was published in 1996 by Drummond et al[8] we will describe the principles for conducting or reviewing an economic analysis with published studies as the example to capture issues particularly relevant to evaluations of EVR and OSR of AAAs, as shown subsequently (Table).

Define a question completely and pose it in an answerable form. Ideally, this should specify the clinical strategies compared, time horizons, measures of clinical effectiveness, target patient populations, and cost-analysis perspective. The question should be asked in a manner that considers both costs and outcomes. The question "Is OSR more costly than EVR of AAAs?" is an incomplete cost-effectiveness question because the decision maker also needs to know the differences in the effectiveness of the two interventions. A more focused question might be "From a societal medical care perspective, is OSR more cost effective than EVR for the initial repair of asymptomatic unruptured AAAs (5 cm in diameter) in providing 2-year survival free of repeat aneurysm-related interventions?" The question should be specific, and the authors should also identify the target population eligible for both interventions (eg, 70-year-old men).

The clinical strategies to be compared, along with measures of their clinical effectiveness and outcomes, should be clearly stated. Examples of clinical consequences for the two strategies include mortality (expressed as crude survival and QALYs), myocardial infarction, pulmonary complications, renal failure, graft infection, wound dehiscence, continued aneurysm sac growth, further rupture, and stroke.

The time horizon in which costs and outcomes are measured in a cost-effectiveness analysis should be clearly stated.[8] The time horizon in the previous example was 2 years. Ideally, the time horizon of an analysis should extend far enough into the future to capture major economic and health outcomes whether they are intended and not (in the example given, to include the impact of later diagnostic procedures, surveillance, and therapeutic interventions).

The target population of interest should be identified. Existing data may or may not be available for a particular population (eg, 70-year-old men with AAA $>$ 5 cm), and with a description of the population of interest, the reviewers are able to judge the validity of the data or perhaps extrapolate the data into a model for their own studies. The perspective (decision-making context) of the cost analysis in the previous example (eg, taking a societal medical care perspective) includes all healthcare costs relevant to patients, payers, and society.

Comprehensively describe the competing alternatives to be measured. True cost-effectiveness analyses are intrinsically comparative, as interventions can only be cost-effective relative to specified alternatives. Given the rate at which vascular procedures evolve, it will be important to clearly describe the operative approaches being compared in a given study of OSR versus EVR. Especially if accumulating experience indicates divergent purchase prices and outcomes for various stents and bifurcated grafts manufactured by various companies for use in EVR, it will be important to distinguish among them when considering costs, frequencies of use, and outcomes of different grafts. Alternately, the decision to aggregate costs or the results of

use of different grafts in the analysis should be acknowledged and justified. The rationale for excluding specific options (eg, watchful waiting for patients with aneurysms > 5 cm) may be widely accepted, but it is important to state that such options were considered but rejected as not appropriate and give the reasons.

Establish the effectiveness of the programs or services. It cannot be emphasized enough that the design of a study influences the amount of bias and confounding present and, therefore, the study's internal validity. Authors should clearly state the design of the study and then describe the data sources used in determining the effectiveness measures or other endpoints of an intervention that were measured to allow the reader to judge the validity of the data and the applicability of the cost-comparison analysis. Aside from performing primary research, investigators can also generate clinical effectiveness data from systematic reviews of clinical trials or formal quantitative syntheses of the literature (metaanalyses).[9,10] In the evaluation of clinical effectiveness for OSR and EVR, the study would first have to be restricted to include a population that was appropriate for both EVR and OSR (eg, to patients without contraindications to general anesthesia) and the success of achieving the stated clinical objective (preventing rupture of AAA) would need to be assessed. The rates of repeat intervention observed during specified postprocedure follow-up should also be tracked. Assessments of perioperative mortality or complications capture procedure safety but are not adequate to measure procedure effectiveness.

Regarding OSR and EVR, initial reports on EVR of AAA claimed that it was a cost-effective alternative to OSR because of the decreased length of stay and the lower rate of postoperative complications.[14-17,22] However, recent questions about its durability and effectiveness over time reflect concerns as to whether or not it is more cost effective than OSR in the long term.[23,24] Previous studies indicated that EVR uses fewer resources and is therefore cost saving relative to OSR during the initial hospitalization period. However, additional interventions from clinical failures have been reported.[24]

The morbidity and mortality rates for OSR of AAAs have been well documented in large randomized controlled or prospective trials.[16,29] However, EVR morbidity and mortality rates have generally been based on hypothetic models with unpublished data (eg, abstract proceedings from scientific meetings) or from consecutive cases from single institutions.[12] A cost-effectiveness analysis (actually a cost-utility analysis) was chosen by Patel et al[12] as the type of economic evaluation to be used to answer the question of whether EVR of AAA is a cost-effective alternative to OSR. The alternatives (OSR and EVR) differ in both costs and effects, and an incremental cost-effectiveness ratio (ICER; to be defined) was calculated to address this question.

Outcome and cost differences must be reasonably ascribable to treatment, rather than other factors. Although uncommon in procedure evaluations,[30] this assessment is best achieved by randomizing patients to the therapies being compared. Nonrandomized studies may suffer from confounding by clinical indication, where clinical features influencing treatment choices also affect outcome, independent of any actual treatment effect.[31] For example, in concurrent intrainstitutional nonrandomized comparative studies of EVR versus OSR, eligible patients presumably underwent EVR. Patients ineligible for EVR (eg, because of unsuitable anatomy) who therefore underwent OSR differed from patients for EVR in ways that potentially affected outcome. This selective referral would therefore confound assessment of treatment effects of EVR versus OSR, and similar processes could complicate interpretation of results of single-arm series of EVR or OSR as well. Ideally, selective use of EVR over OSR in institutions where both procedures are concurrently available should be explicitly addressed. Analyses on the basis of literature data may be forced to include consecutive series (especially of EVR) from papers that do not describe the criteria used to select patients in whom EVR was performed.

Given the possibility that outcome differences could reflect dissimilarities in patient characteristics rather than in treatment effectiveness, treatment group comparability in relevant parameters (eg, gender, age, specific comorbidities affecting operative risk) is essential. This should include specifying criteria (eg, threshold of hypertension, evidence of diabetes treated with oral agents or insulin, current versus any history of tobacco use) used to establish the presence of risk factors. However, the mere absences of statistically significant differences in measured factors do not prove treatment group similarity, especially in studies with limited statistical power.[32] Also, unless measured variables capture key patient characteristics (eg, aneurysm morphology, systemic risk factors) influencing original treatment choice and outcome, any analysis of risk factors is of limited value in showing prognostic similarities in the groups being compared. Alternatively, studies can adjust for residual confounding with various analytic approaches, including logistic regression or propensity scoring.[33]

Even nonrandomized comparisons should follow the intention-to-treat principle commonly used to analyze results of randomized controlled trials.[34] The primary analysis should ascribe costs and outcomes to the original treatment group for patients in whom any AAA repair attempt is started but then converted to another procedure. For example, it is not appropriate to exclude patients in whom EVR was unsuccessful or converted to OSR or to ascribe their results (successes or failures) to the OSR ultimately performed. Results on the basis of final treatment or those restricted to costs of successful interventions may also be examined, but such analyses should clearly be secondary.

Identify important and relevant costs and consequences for each alternative. EVR studies should capture all costs consistent with the specified analytic perspective, including those of relevant diagnostic tests. It is reasonable to either include or exclude tests common to both procedures, as long as this is done consistently. Usage and costs should be included for preoperative diagnostic tests preferentially performed with one treatment method

or the other (eg, duplex ultrasound scan, computed tomographic [CT] scanning with/without contrast, arteriography, or spiral CT scanning). A base case scenario should include costs of more extensive follow-up (eg, more frequent CT scans or duplex scans in the first year and then yearly) for EVR because postoperative surveillance is a major additional cost for this procedure. Hospital perspective analyses of healthcare costs should estimate institutional costs, including graft and other material purchase prices, equipment purchase, personnel costs, and hospital overhead.[9] Although studies that take the hospital perspective would properly exclude costs of care provided by nonsalaried physicians or physician assistants, such studies should include costs of these personnel if they are hospital employees.

Societal perspective analyses of healthcare costs of EVR versus OSR should include hospital costs, as noted previously, and costs of observed involvement of multiple specialists (ie, cardiac or vascular surgeons, anesthesiologists, radiologists, and other physicians) and of physician-employed or hospital-employed physician assistants if applicable. Operating room times that differ for EVR and OSR would generate cost differences reflecting variable duration of anesthesiologist or surgeon involvement and higher surgeon reimbursement for EVR relative to OSR, such that surgeon costs (captured by reimbursement of some other approach) should be included. Exclusion of costs that are similar for both procedures (eg, surgeon reimbursement if it is the same for both procedures under some payment plans) might not affect net cost differences from a societal perspective. However, if frequencies of repeat procedures or treated complications differ, ignoring surgeon costs will distort treatment strategy costs.

It is now recommended that cost analyses take a broader societal perspective that includes capture of patient out-of-pocket healthcare and related costs, nonmedical economic costs of lost productivity (eg, lost wages), and costs that healthcare providers incur when caring for patients.[9] Aside from citing the literature source (including specifying the use of local versus national averages of Medicare physician reimbursement rates) of external cost data used, authors should clearly indicate the nature of costs being included (eg, including versus excluding postdischarge nursing home care when considering prolonged convalescence after AAA repair).

The study design should also reasonably capture key clinical outcomes of interest. In studies of EVR versus OSR, clinical consequences of interest for both procedures would include mortality, myocardial infarction, pulmonary complications, renal failure, continued aneurysm sac growth to some specified diameter, aneurysm rupture, and stroke.[24] For EVR, consequences requiring intervention would include graft thrombosis or occlusions requiring thrombolysis or reoperation. Endoleaks are of importance if they require additional interventions such as stenting, coil placements, or conversion to OSR. Postoperative interventions of interest after OSR include reoperation for hemorrhage or dehiscence, recurrent aneurysms, and graft infection.

It is important to establish consistent criteria for the duration of short-term (eg, within 30 days after surgery) versus longer term (eg, 2 years or more) follow-up versus lifetime follow-up. Authors should clarify methods used to address censoring (eg, total lifetime AAA repair costs or outcomes being unknown for patients still alive or lost to follow-up at the end of the study), which can have various effects on frequencies of outcomes and on cost estimates.[34] When modeling long-term outcomes, it is also important to distinguish between observed follow-up and extrapolated estimates on the basis of modeling.[9,27]

One situation potentially complicating tracking of costs and outcomes over time is that patients may spend time in various health states or levels of wellness such as being well, sick, or dead, with attendant costs associated with each state. Although conventional decision trees just consider the long-term probability of attaining each state, an approach called Markov modeling allows consideration of probabilities of transitioning among states during short time intervals (Markov cycles)[9] and then aggregating costs or quality-adjusted survival associated with the total time spent in each state.[8,9,27]

Costs and consequences should be measured accurately in appropriate units. To allow readers to better estimate costs for their settings, resource use should also be measured in appropriate units (eg, length of operating room stays in hours or total hospital stays in days) and presented separately from costs. Resource use is then multiplied by reasonable estimates of costs per unit of resource use to generate costs for individual care components. Currency is the best summary measure of aggregate resource use costs. Estimates of costs for care provided in different calendar years (especially those from the literature) should be expressed in constant year (eg, 2002) dollars with the Medical Care component of the local or regional Consumer Price Index.[35] Clarity of methods and results would be enhanced by also expressing costs in dollars from the original year in which the data were collected.

Outcomes of interest would include proportions of patients with AAA repairs successful in excluding the aneurysmal section of aorta, as measured with intraprocedural imaging or with freedom from subsequent rupture or repeat interventions. Success of the procedure is a separate issue from the occurrence of certain complications (untoward consequences, such as a myocardial infarction), which may or may not occur, independent of the success of the procedure. Major complications (eg, renal failure) should be distinguished from minor complications (eg, superficial wound infections). Complications relevant to AAA repair could include all-cause mortality, myocardial infarction, pulmonary complications, limb ischemia requiring major amputation, renal failure, graft infection, wound dehiscence, continued AAA sac growth, rupture, and stroke.[12-25] It is important to note the duration of follow-up and the cost components included (eg, inclusion of inpatient and outpatient facility and physician care costs, nursing home costs, costs of lost productivity). Studies also need to identify specific outcomes (eg, one or more key

primary outcomes or a composite such as 2-year survival free of rupture or of repeat intervention) for use in calculating cost-effectiveness ratios (described subsequently).

The costs and consequences should be valued credibly according to the stated cost analysis perspective. Studies reporting hospital costs need to specify cost estimate sources (eg, true production costs, hospital-wide or department-specific cost to charge ratios).[36] Costs of usage of C-arm fluoroscopy and intravascular ultrasound scan should be based on observations or clearly articulated assumptions regarding amortized equipment purchase prices and related technologist and radiologist time. Although detailed activity-based cost analysis of the use of diagnostic equipment, technologists, and radiologists[37,38] may be beyond the scope of a given study, reasonable and unbiased estimates of costs associated with such imaging should be made.

Studies measuring societal perspective medical care costs should combine hospital production costs or reimbursement with physician costs associated with specific treatment-related tasks, as estimated with Resource Based Relative Value Scales[39] (reflecting the quantity and intensity of physician effort expended in providing treatment), observed time spent in patient care multiplied by compensation and practice expenses expressed per unit time, or reimbursement (eg, Medicare fee schedules). Billed charges do not adequately represent costs, as charges represent neither true costs of providing care nor the reimbursement actually provided by insurers or patients.[36] If charges are used because of the absence of better data, then the deficiencies of charges as a measure of cost should be clearly acknowledged.

Measured outcomes should also have been captured comprehensively. For example, analyses capturing postdischarge outcomes should include admissions to other facilities (eg, via review of referring physician office records).

Costs and outcomes should be adjusted for differential in timing. If the time horizon extends beyond the interval for which primary outcomes data are available, then modeled data must be used in the analysis.[27] Initial cost-effectiveness analyses comparing OSR with EVR of AAAs have used only short-term time horizons up to this point in time because there were no long-term outcomes data (primary data) available or the authors were only interested in determining the differences in length of hospitalization and costs of initial care.[13-15,17-19] More recently, cost-effectiveness analyses with a long-term time horizon incorporating modeled data have been conducted.[12] If the analysis is not extended far enough in the future to capture important differences in cost or outcomes, then short-term results may not accurately reflect eventual outcomes.[8]

Costs observed beyond 1 year of follow-up for individual patients should be discounted to present value (eg, at 3%/y).[8,9] This is independent of inflation and reflects[9] potential interest or investment income associated with receiving a given dollar amount 2 years from now rather than today. For studies following individual patients less than 1 year, discounting is deemed not applicable. For studies with longer follow-up, the discount rate should be stated and the reason for choosing a particular rate should be justified. Although costs are generally always discounted, discounting of health outcomes is controversial; however, an explanation should be provided if costs and outcomes are not discounted.[9] A discussion of the controversies surrounding discounting of treatment effects is beyond the scope of this paper.

An incremental analysis of costs and consequences of alternatives should be performed. Ideally, studies should measure both costs and clinical effectiveness. Those finding one treatment to be more expensive but more effective should present ICERs that quantify net cost increases relative to net improvements in specified outcomes. For example, studies finding EVR to generate higher 1-year survival free of the need for AAA repair at higher average cost than OSR could express these results as mean incremental costs per added year of survival free of untoward events (including major complications, reinterventions). This is illustrated in the following sample calculation on the basis of factitious data.

Where:

Cost_{EVR} = mean lifetime cost per EVR patient (including costs of future care) = \$27,000

Cost_{OSR} = mean lifetime cost per OSR patient (including costs of future care) = \$20,000

Eff_{EVR} = mean (modeled) lifetime effectiveness of the EVR strategy = 7.8 QALYs

Eff_{OSR} = mean (modeled) lifetime effectiveness of the OSR strategy = 7.5 QALYs

Then:

$$\frac{\text{Cost}_{\text{EVR}} - \text{Cost}_{\text{OSR}}}{\text{Eff}_{\text{EVR}} - \text{Eff}_{\text{OSR}}} = \frac{\$27{,}000 - \$20{,}000}{7.8 - 7.5 \text{ QALYs}} = \frac{\$7{,}000}{0.30 \text{ QALYs}} = \$23{,}333/\text{QALY}$$

This result indicates that, relative to OSR, EVR has an ICER of \$23,333/QALY. This means that each additional \$23,333 spent with EVR rather than OSR should "buy" 1.0 additional QALYs. The interpretation of such a ratio (eg, whether or not it is considered to represent good "value for money") should be addressed in the discussion (see The presentation and discussion of results should include issues that are of most concern to potential users).

One procedure could be superior to the alternative in either the cost or outcome dimension and also be at least as good, if not better, in the other dimension. In such cases, the superior procedure dominates the alternative. Dominance is relatively unusual but, when present, is an appropriate justification for not calculating ICERs.[8,9,27]

Allowance should be made for uncertainty in the estimates of cost and consequences. Levels of uncertainty are associated with cost-effectiveness analyses because of the multiple pieces of information being combined (natural history of the disease, effectiveness and costs of interventions, preferences regarding health outcomes,

long-term results of interventions). To reach defensible study conclusions, authors must assume that the true values for all parameters are known or that available figures are unbiased estimates of the true value and that the characteristics of the population are known. Authors should report the uncertainty associated with any observed costs or outcome differences.[9]

For stochastic data, details of statistical tests performed (including assumptions regarding normal distributions) should be provided along with confidence intervals for point estimates of individual results and calculated ICERs.[40-42] Appropriate measures of central tendency (eg, means versus medians) of costs and confidence intervals for outcomes frequencies (eg, mortality, graft infection) need to be provided. If applicable, an analysis of costs and outcomes adjusted for differences in risk factors should be performed. A distinction should be made between nonsignificant results that are clinically or economically important but reflect small sample size and those that reflect significant differences that are clinically or economically unimportant.

To assess the effects of the various assumptions made in a cost-effectiveness analysis, a sensitivity analysis should be performed.[8,9,27,40,42] Different approaches to sensitivity analyses include one-way or univariate sensitivity analyses and multiway or multivariate sensitivity analyses. One-way sensitivity analyses assess how changing values of single parameters within ranges consistent with available clinical or economic data alter the base case ICER.[9] An analysis is said to be sensitive to values of a particular variable if changing its value (consistent with the data at hand) affects the conclusion of the analysis (eg, which treatment is favored). Conversely, the analysis is said to be robust or insensitive to values of that variable if changes within a reasonable range do not alter the conclusion.[42]

In a multivariate sensitivity analysis, the expected outcome is determined for different combinations of estimates of the values of either two or three variables (two-way or three-way sensitivity analysis, respectively) while holding all other variables at their base case values. If one-way or multiway sensitivity analyses indicate that conclusions are sensitive to reasonable changes in key parameter values, then authors should be appropriately circumspect about the strength of their conclusions and may consider seeking better data to reduce uncertainty about the values of those parameters. Conversely, having results that are robust over reasonable ranges of multiple parameters increases one's confidence in the study conclusions, regardless of the value of the individual parameters. Calculating confidence intervals for ICERs should also be done to increase the confidence in the study conclusion.

The presentation and discussion of results should include issues that are of most concern to potential users. It is important to answer all components of the cost comparison study question (eg, "From a societal medical care perspective, is OSR more cost-effective than EVR for the initial repair of asymptomatic unruptured AAAs (5 cm in diameter) in providing 2-year survival free of repeat aneurysm-related interventions?"). Although no study can address all concerns of interest to all readers, the discussion should be reasonably comprehensive. Conclusions should be consistent with data-driven results and the proper terms such as cost-effective[ness] should be used. Assertions of cost-effectiveness require support by data on clinical effectiveness and cost. This includes a discussion as to whether or not the ICER is sufficiently low as to be consistent with those of other interventions that society considers cost effective. Procedures do not have to be dominant or cost saving to be cost effective. The threshold value of $60,000/QALY quoted by Patel et al[12] reflects a widely used convention. ICERs below such a value favor classifying the procedure of interest as cost effective and therefore worthy of dissemination. Conversely, ICER values above such thresholds may identify treatments with effects that do not justify the costs incurred. Because this particular threshold value in part reflects conversion into current dollars of the classic ICER threshold used to justify Federal funding of dialysis in the End Stage Renal Disease program begun 25 years ago, it is admittedly arbitrary and may not be an appropriate cutoff for every situation.

Ideally, discussions should also compare the study results with those of previously published reports, acknowledging similarities and differences in study design, patient population, cost perspective, and outcomes. Studies should also consider the generalizability and potential policy implications of their results. For example, authors studying repair of AAAs greater than or equal to 5 cm in diameter might be tempted to apply their findings to smaller aneurysms (AAA less than 5 cm) as well. Any discussion of the implications of such an assumption should acknowledge the risks of extrapolating results to a larger or different population that was not studied directly and may differ in gender distribution, operative risks, and other features. The authors should also discuss the broader implication of their study results (eg, implications for decision makers with other perspectives).

Most importantly, the authors need to acknowledge the study limitations and their effect on the study conclusion (eg, using charge data to measure costs, comparing different populations, using data from unpublished studies, conducting a one-way sensitivity analysis instead of a multivariate one). As mentioned earlier, the design of a study influences the amount of bias and confounding present and, therefore, a study's internal validity, so it is paramount that the authors consider the design of the study in which they abstracted effectiveness data when discussing the results.

CONCLUSION

Given the increased awareness of the importance of cost effectiveness in healthcare, there has been a growth in the number of published economic evaluations in recent years. In response to concerns about deficiencies in the methodology of published studies, there has been a concerted effort to use established guidelines or standards when conducting or reviewing economic evaluations. This paper has

briefly defined the various forms of economic evaluations and used a clinical example familiar to the readers of this Journal to describe some useful steps for conducting an appraisal of a cost-effectiveness analysis. Our intent was not to actually conduct such an analysis with existing data but to describe the principles for conducting or reviewing an economic analysis with realistic data from the literature. We used a well-known checklist that was published in 1996 by Drummond et al[8] that has been widely used by authors and editors of several journals (eg, British Medical Journal, Journal of the American Medical Association, New England Journal of Medicine, Health Economics).[43-46] The guidelines or policies for conducting cost comparisons that have been published dictate that investigators document the degree to which appropriate methodology supports their results.

REFERENCES

1. Hallet JW Jr, Marshall DM, Petterson TM, Gray DT, Bower TC, Cherry KJ Jr, et al. Graft-related complications after abdominal aortic aneurysm repair: reassurance from a 36-year population-based experience. J Vasc Surg 1997;25:277-84.
2. Stainer CA, Powe NR, Anderson GF, Das A. Technology coverage decisions by health care plans and considerations by medical directors. Med Care 1997;35:472-87.
3. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analyses in randomized control trials. Med Care 1992;30:231-43.
4. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med 1977;296:716-21.
5. Eisenberg JM. Clinical economics. A guide to the economic analysis of clinical practices. JAMA 1989;262:2879-86.
6. Detsky AS, Naglie IG. A clinician's guide to cost effectiveness analysis. Ann Intern Med 1990;113:147-54.
7. Gray DT. Cost effectiveness analysis of interventions for congenital heart disease. J Interv Cardiol 1995;8:579-7.
8. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. Methods for the economic evaluation of health programmes. 2nd ed. Oxford: Oxford University Press; 1996.
9. Gold MR, Russell LB, Siegel JE, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.
10. Mugford M. Using systematic reviews for economic evaluation. In: Egger M, Altman D, Davey-Smith G, editors. Systematic reviews. 2nd ed. London: BMJ Books; 2001. p. 419-28.
11. Gray DT, Veenstra DL. Comparative economic analyses of minimally invasive direct coronary artery bypass surgery. J Thorac Cardiovasc Surg. In press.
12. Patel ST, Haser PB, Bush HL, Kent KC. The cost-effectiveness of endovascular repair versus open surgical repair of abdominal aortic aneurysms: a decision model. J Vasc Surg 1999;29:958-72.
13. Clair DG, Gray B, O'Hara PJ, Ouriel K. An evaluation of the costs to health care institutions of endovascular aortic aneurysm. J Vasc Surg 2000;32:148-52.
14. Stenbergh WC III, Money SR. Hospital cost of endovascular versus open repair of abdominal aortic aneurysms: a multicenter study. J Vasc Surg 2000;31:237-44.
15. Seiwert AJ, Wolfe J, Whalen RC, Pigott JP, Kritpracha B, Beebe HG. Cost comparison of aortic aneurysm endograft exclusion versus open surgical repair. Am J Surg 1999;178:117-20.
16. Schermerhorn ML, Birkmeyer JD, Gould DA, Cronenwett JL. Cost-effectiveness of surgery for small abdominal aortic aneurysms on the basis of data from the United Kingdom small aneurysm trial. J Vasc Surg 2000;31:217-26.
17. Bosch JL, Lester JS, McMahon PM, Beinfeld MT, Halpern EF, Kaufman JA, et al. Hospital costs for elective endovascular and surgical repairs of infrarenal abdominal aortic aneurysms. Radiology 2001;220:492-7.
18. Liapis C, Kakisis J, Kaperonis E, Papavassiliou V, Karousos D, Tzonou A, et al. Changes of the infrarenal aortic segment after conventional abdominal aortic aneurysm repair. Eur J Vasc Endovasc Surg 2000;19:643-7.
19. Prinssen M, Buskens E, Blankensteijn JD. The Dutch randomized endovascular aneurysm management (DREAM) trial. Background, design and methods. J Cardiovasc Surg (Torino) 2002;43:370-84.
20. Ceelen W, Sonneville T, Randon C, DeRoose J, Vermassen F. Cost-benefit analysis of endovascular versus open abdominal aortic aneurysm treatment. Acta Chir Belg 1999;99:64-7.
21. Quinones-Baldrich WJ. Achieving cost-effective endoluminal aneurysm repair. Semin Vasc Surg 1999;12:220-5.
22. Birch SE Stary DR, Scott AR. Cost of endovascular versus open surgical repair of abdominal aortic aneurysms. Aust N Z J Surg 2000;70:660-6.
23. Holzenbein J, Kretschmer G, Glanzl R, Schon A, Thurnher S, Winkelbauer F, et al. Endovascular AAA treatment: expensive prestige or economic alternative? Eur J Vasc Endovasc Surg 1997;14:265-72.
24. Dattilo JB, Brewster DC, Fan CM, Geller SC, Lauterbach SR, Abbott WM. Clinical failures of endovascular abdominal aortic aneurysm repair: incidence, causes, and management. J Vasc Surg 2002;35:1137-44.
25. Patel ST, Korn P, Haser PB, Bush HL Jr, Ken KC. The cost-effectiveness of repairing ruptured abdominal aortic aneurysms. J Vasc Surg 2000;32:247-57.
26. Marshall JK, Cawdron R, Yamamura DLR, Ganguli S, Lad R, O'Brien BJ. Use and misuse of cost-effectiveness terminology in the gastroenterology literature: a systematic review. Am J Gastroenterol 2002;97:172-9.
27. Petitti DB, editor. Meta-analysis, decision-analysis, and cost-effectiveness analysis. 2nd ed. New York: Oxford University Press; 2000.
28. Mauskopf JA, Paul JE, Grant DM, Stergachis A. The role of cost-consequence analysis in healthcare decision making. Pharmacoeconomics 1998;13:277-88.
29. Johnston KW, Canadian Society for Vascular Surgery Aneurysm Study Group. Non-ruptured abdominal aortic aneurysm: six year follow-up results from the multicenter prospective Canadian aneurysm study. J Vasc Surg 1994;20:163-70.
30. Gray DT, Hewitt P, Chalmers TC. The evaluation of surgical therapy. In: Rutkow IM, editor. The socioeconomics of surgical health care delivery. St Louis (MO): CV Mosby Co; 1989. p. 228-56.
31. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. Int J Epidemiol 1980;9:361-7.
32. Blackstone EH. Comparing apples and oranges. J Thorac Cardiovasc Surg 2002;123:8-15.
33. Fisher L, Dixon D, Herson J, Frankowski R, Hearon M, Peace K. Intention to treat in clinical trials. In: Peace K, editor. Statistical issues in drug research and development. Statistics: textbooks and monographs, vol 106. New York: Marcel Dekker, Inc; 1990. p. 331-50.
34. Etzioni RD, Feuer EJ, Sullivn SD, Lin D, Hu C, Ramsey SD. On the use of survival analysis techniques to estimate medical care costs. J Health Econ 1999;18:365-80.
35. Bureau of Labor Statistics. CPI (CPI-U): US city average. Washington (DC): US Department of Labor; 2002.
36. Shwartz M, Young DW, Siegrist R. The ratio of costs to charges: how good a basis for estimating costs? Inquiry 1995;32:476-81.
37. Rubin GD, Amerding MD, Dake MD, Napel S. Cost identification of abdominal aortic aneurysm imaging by using time and motion analyses. Radiology 2001;215:63-70.
38. Gray DT, Hollingworth W, Blackmore CC, Alotis MA, Martin BI, Sullivan SD, et al. Conventional radiography, rapid magnetic resonance and conventional magnetic resonance imaging for low back pain: activity-based costs and reimbursement. Radiology. Accepted.
39. Hsiao WC, Couch NP, Causino N, Becker ER, Ketcham TR, Verrilli DK. Resource-based relative values for invasive procedures performed by eight surgical specialties. JAMA 1988;260:2418-24.
40. Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Wein-

stein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996. p. 247-75.
41. Polsky D, Glick HA, Willkie R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Econ 1997;6:243-52.
42. Briggs A, Gray A. Handling uncertainty when performing economic evaluation of health care interventions. Health Technol Assess 1999;3: 1-132.
43. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ 1996;313:275-83.
44. Kassirer JP, Angel M. The journal's policy on cost effectiveness analyses. N Engl J Med 1994;331:669-70.
45. Schulman K, Sulmasy DP, Roney D. Ethics, economics and the publication policies of major medical journals. JAMA 1994;272:154-6.
46. Jefferson TO, Demicheli V. Are guidelines for peer reviewing economic evaluations necessary? A survey of current editorial practice. Health Econ 1995;4:383-8.

Submitted Aug 13, 2002; accepted Sep 13, 2002.

Additional material for this article may be found online at www.mosby.com/jvs.